TECHNICAL REPORT NO. LWL-CR-03B70

DETECTION OF EMOTIONAL STRESS BY VOICE ANALYSIS

Final Report

By

Decision Control Incorporated
4809 Auburn Avenue
Bethesda, Maryland 20014

September 1972

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

U. S. ARMY LAND WARFARE LABORATORY
Aberdeen Proving Ground, Maryland 21005

TECHNICAL REPORT NO. LWL-CR-03B70

DETECTION OF EMOTIONAL STRESS BY VOICE ANALYSIS

Final Report

By

Decision Control Incorporated
4809 Auburn Avenue
Bethesda, Maryland 20014

September 1972

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

U. S. ARMY LAND WARFARE LABORATORY
Aberdeen Proving Ground, Maryland 21005

ABSTRACT

A study was performed by Decision Control Incorporated (DCI) to develop a methodology for determining emotional stress by voice analysis. Tape recordings of 58 subject examinations for actual or staged criminal offenses were electronically processed. The result of this processing was a measure value for the ratio of the amplitudes of the 100-120 Hz to the 600-800 Hz ranges in each subject's spoken response. A methodology for assessing the measure values to determine the relative levels of stress in these responses was developed. A portable prototype voice analyzer was built which electronically processes a voice signal from either a tape recorder or microphone input. An interrogation and assessment protocol similar to the "Zone of Comparison" technique used by many polygraph specialists was developed and used to assess the guilt or innocence of five subjects undergoing "Zone of Comparison" tests. A 95% confidence level was achieved for four of the five subjects.

FOREWORD

A study with a twofold purpose was performed by Decision Control Incorporated (DCI). The study was to develop a methodology for determining emotional stress by voice analysis. The second objective was to modify a breadboard voice analyzer developed under Contract No. DAAD05-70-C-0176 to that extent necessary to make it suitable for field applications.

Previous studies by DCI had shown that an amplitude/frequency analysis of the response "no" would provide a fairly accurate assessment of whether the response was truthful or deceitful (approximately 70 percent accuracy). It was desirable to establish the accuracy of voice analysis as a means of detecting emotional stress and to develop a protocol for the use of the voice analyzer which achieves the highest possible accuracy.

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

I. INTRODUCTION

Present methods of lie detection rely on the assumption that when an individual is asked a question and his response to that question is deceitful, that individual enters a stressful psychological state which is evidenced by certain physiological irregularities. In present polygraph technique, the interrogated subject is asked a series of questions and instructed to respond "yes" or "no" to each. During the entire period of the test the subject is physically connected to a polygraph and three parameters of his physiological state are monitored; the blood pressure and pulse, the rate of respiration, and the electrical resistance of the skin. The measurements of each are known as the cardiograph, the pneumograph, and the galvanic skin response (GSR) respectively. While various polygraph experts disagree over which measure provides a more reliable and accurate indication of psychological state, all three are gross indicators. DCI and others feel that a more precise physiological parameter, which is more closely associated with the cognitive processes should provide a better indication of psychological state. Such a parameter is speech. The speech process is controlled by a large number of extremely sensitive and quickly acting muscles. In a highly complex manner, these muscles react to both the cognitive and emotional processes of an individual.

Previous studies[1] have resulted in strong evidence that there exists the possibility of detecting accurately a stressful psychological state in an individual by analyzing that individual's speech during such a state. Further, two previous studies by DCI[2] have shown that an analysis of the response "no" could provide a fairly accurate assessment of deceit on the part of an individual. This assessment was based on the observation that when an individual responded to a question with a known lie, there typically occurred a shift in the total amount of energy present in certain frequency ranges of the individual's speech signal. For the sample population, the maximum changes occurred around the male speaker's fundamental frequency of 100 to 120

---

[1] Experimental Investigation of Voice Changes in Lying; Fred H. Fuller; Proceedings, 2nd National Symposium on Law Enforcement Science and Technology, Chicago, Illinois, 1968

Frequency Spectrum of Speech as an Indicator of the Degree and Nature of Emotional Stress; V.A. Popov, P.V. Simonov, et.al.; JPRS 52698; 23 March 1971

A Potential New Measurement of Emotional State; E.N. Whitman and D.J. Flicker; Journal of Newark Beth Israel Hospital; Vol. XVII, No. 3.; pp 167-8, July 1966

[2] Validation Program for Lie Detection Techniques Using Voice Analysis

cycles per second and the mid-range frequencies of 500 to 800 cps. It was not within the scope of these studies to determine the precise psycho-physiological causes of this shift. It was sufficient that they were highly correlated and subject to replication.

Detecting and quantifying these phenomena has been accomplished using a voice parameter consisting of the ratio of the two affected frequency ranges. A breadboard analyzer was developed which could automatically extract and measure this parameter.

August 1969, Purchase Order No. DAAD05-69-M-5025
Application of Voice Analysis Method, February 1971, Contract Number DAAD05-70-C-0176

II. CONCLUSIONS

1. Based on a number of independent tests and procedures, the ratio of the peak amplitudes of the 100-120 Hz band and the 600-800 Hz band of frequencies of the human voice provides a reliable measure of emotional stress and/or deceit.

2. Based on the responses for the ratio measure value using a standard stimulation test, an acceptable level of assessment accuracy is achieved by eliminating those responses whose magnitude deviates less than a given value from the mean response level, and are thereby not amenable to accurate assessment. (This typically encompasses the mid-range of any rank ordering of a set of responses). When approximately 30% of the responses having the smallest deviation from the mean response level are eliminated, the assessment accuracy for the remaining responses is at least 70%.

3. The prototype voice stress analyzer, when used with the zone of comparison protocol provides an effective means of assessing guilt or innocence with high confidence. Further testing of at least 25 to 50 additional subjects should be accomplished, however, to properly validate and extend the results presented in this report.

III. DISCUSSION OF SUBJECTS TESTED

The investigation for the present project consisted of the analysis of numerous tape recordings of subject interrogations for actual and staged crimes. These interrogations can be grouped into two major categories.

Group A consists of those individuals who underwent what is termed a "peak-of-tension" test. This test is typically designed such that a subject is asked the same question concerning a number of similar items and is expected to react to only one of these questions as evidenced on a polygraph. In a "card stimulation test" the subject is asked to pick one of six cards, numbered one through six. He would then be asked if the number he picked was one, two, etc. The subject would be previously instructed to respond with a "no" to each question. Hence, there is only one deceitful and theoretically stressful response in the test. Similarly, a subject under interrogation for an actual or staged crime might be asked a set of questions relating to a number of similar items, one of which was directly related to the crime. This test is defined as a peak of tension test. If the subject were guilty or knowledgeable of the crime, he would be expected to react to the crime-specific item. Such a test might consist of the question, "Was the gun used a Browning", a "Luger", etc.

Under Group A, Group A-1 consisted of nineteen subjects undergoing a card stimulation test conducted in Hebrew or Arabic. This special group was analyzed to determine the validity of the technique for a non-Romance foreign language. Group A-2 consisted of 13 subjects undergoing the card stimulation test. Group A-3 consisted of 6 subjects who underwent examinations for staged crimes, using the peak of tension test.

The second major category, Group B, consisted of a number of individuals who underwent a test referred to as the "Zone of Comparison Test". Several variations of this test were encountered in this project. In Group B-1, the subject would first be asked several innocuous questions. He would then be asked, in an alternating sequence, questions which relate to the offense in question, and questions totally unrelated to the crime, which are expected to evoke a stressful response. An example of this last type of question might be: "Did you ever steal money out of your mother's purse?".

Group B-2 used alternating crime-related and innocuous questions. Group B-3 consisted of 9 subjects undergoing a Zone of Comparison Test for a staged crime. The subjects could either be innocent, have compromised classified material or have stolen classified material. These subjects'

tests included six crime-related questions, three of which related to the compromising, the other three related to the theft of the documents. The last question in each set of questions was non-crime-related, but considered to be stress-inducing. The question specifically asked was: "While in high school, did you ever cheat on an exam?". It was assumed that each subject would reply deceitfully to this last question.

The following table summarizes the tests described above:

CATEGORY	TYPE OF QUESTION	TYPE OF TEST
Group A-1	Did you pick card number ___?	Card Stimulation
Group A-2	Did you pick number ___?	Card Stimulation
Group A-2	Was the gun used a ___?	Peak of Tension
Group B-1	Are you sometimes called Dick? Did you steal that gun? Did you ever steal from your mother?	Zone of comparison
Group B-2	Are you wearing blue pants? Did you steal that camera?	Zone of comparison (modified)
Group B-3	Are you ready to begin? Did you memorize the contents? Did you steal that document?	Zone of Comparison

Table 1
Tests and Typical Questions Encountered in Study

IV. DISCUSSION OF ANALYSES PERFORMED

The tape recordings were played back on an Ampex AG-600 tape recorder at the same speed at which they were recorded. The signal output of this tape recorder was filtered through a variable bandpass filter (Krohn-Hite 310-A), rectified, smoothed, and recorded on a graphical analog recorder (E and M Physiograph Six) for Group A and for Group B-1 subjects. Tape recordings for Group B-2 and B-3 were processed with the prototype voice analyzer, and are discussed later.

The tape recordings were processed for a number of bandpass frequencies. The voice responses of Group A-1 were processed for the following bandpass frequencies:

100-120Hz	200-300Hz	800-1000Hz	3000-5000Hz
120-140	300-400	1000-1200	5000-7000
140-160	400-600	1200-1500	7000-10,000
160-200	600-800	2000-3000	100-10,000

The output on the graphic recorder was manually reduced to a measure of peak amplitude of response for each bandpass frequency. These response measure values were graphically displayed on barcharts and rectilinear coordinate graphs. A sensitivity analysis was then made for each bandpass. This analysis consisted of the development of 'hit-miss' ratios for each bandpass and additive combinations of bandpasses. A 'hit-miss' ratio consists of the number of times the protocol used successfully discriminates deceit to the number of times the assessment was incorrect. Results indicated that the 100-120 Hz and 600-800 Hz bandpass regions were the best groups and that their values changed in opposite directions under conditions of stress. The use of the 100-10,000 Hz bandpass, termed all-pass, presented some possibilities. Hence, all subsequent analyses used the 100-120 Hz and 600-800 Hz frequency bands except when the low bandpass (100-120 Hz) signal-to-noise ratio was too low. In these cases, the high bandpass (600-800 Hz) and the all-pass values were used.

The ratio values, low bandpass/high bandpass and high bandpass/all bandpass were also developed and graphically displayed for each response. These charts and graphs were assessed and conclusions drawn as to the protocol most suitable for the identification of stress (or deceit) based on these charts.

The measure values developed were ranked within each subgroup. Those responses which had been identified as stressful by the responsible polygraph personnel or as lies by the interrogators were then plotted as a cumulative (and density) distribution or a function of the ratio value, on normal probability graph paper.

The results show that the data are approximately normally distributed. The remaining non-stressful (truthful) responses of each subgroup were similarly plotted on the same graph. An intermediate curve was then developed which represented the probability of an accurate assessment as a function of the measure value. This intermediate curve was used to generate the final curve, which showed the accuracy of the assessment versus the percentage of the total responses which had been assessed for each subgroup. For simplicity, this analysis assumes an equal number of stressful/deceitful and non-stressful/truthful responses.

An additional procedure was undertaken to further evaluate the measured responses. This was the rank-ordering by subject of the ratio values of the responses within each test. It was felt that the stressful responses should be the lowest ranked or highest ranked values, depending on which ratio or bandpass was examined. The same statistical analysis was carried out for this procedure as for the previous one.

The use of the accuracy of assessment versus the percentage of responses assessed requires further discussion. The results of the preliminary data analysis indicated that there exists a grey area in which an assessment of veracity or stress cannot be made with any reasonable accuracy. This grey area is the middle range of the measure values of the responses from the sample tested. An example should clarify this. Assume a subject has made ten voice responses in a test and measure values have been obtained from these responses. Let these ten measure values be rank-ordered from one to ten, low to high. Based on the developed protocol, the lowest measure value response, rank-ordered one, would be assessed as stressful. If this assessment was correct, based on validating polygraph information, an accuracy of 100 percent is realized, having assessed 10 percent of the responses. If the highest measure value response, rank-ordered 10, were correctly assessed, then 20 percent of the responses would now have been assessed with an accuracy of 100 percent. Following this procedure, the three lowest ranked and four highest ranked responses might be assessed as stressful and nonstressful, respectively. Seventy percent of the responses have now been assessed. If one of those assessments was incorrect, an accuracy of 6/7 or 86 percent has been realized. The responses rank-ordered 4, 5, and 6 have not been assessed. This is the grey area.

V. RESULTS OF ANALYSES

The voice data were processed to obtain the desired measure values and ratio values. Meetings were held with the polygraph specialists responsible for the evaluation of the polygraph data to obtain their assessment of the responses. For those tests where only an assessment of Deception Indicated (DI) or No Deception Indicated (NDI) was available from polygraph examinations, this assessment was used as the basis for analysis. For those tests where no polygraph assessment was available, the known lies were used as a basis for analysis. The following sections discuss the results obtained for the various subgroups.

1. GROUP A-1

This group consisted of 19 subjects speaking Hebrew or Arabic. The test consisted of each subject picking one of six numbered cards. When the interrogator asked the subject if the card he picked was card number one, number two, and so on, the subject responded "no" to each including the one he picked. There were two repetitions of the six questions for each subject, thus there were two deceitful responses from each subject. No polygraph charts were available for this group.

The results of the sensitivity analysis have been discussed previously. Analysis of the measure value data indicated the presence of an unexpected phenomenon. The ratio measure value did indeed decrease for many of the subjects when they lied, yet for a number of subjects it tended to increase. This indicated that while for about two-thirds of the subjects a deceitful response was indicated by a shift in energy from the low frequency range to the high range, a reversal of this phenomenon was observed in about a quarter of the cases. In order to discriminate between the two "opposite" types the measure value of the first response of each subject was examined. A high positive correlation was observed to exist between the first response and the deceitful response relative to the rank order of each within the majority of the subject's test groups.

Each subject's test group was then re-examined with the following logic: if the first response was low relative to the other responses, then the lowest response (excluding the first response) would be assessed as deceitful. Similarly, if the first response was observed to be high, then the highest response was assessed as deceitful. If the first response was neither low nor high, then the lowest response was suspect.

The results of the statistical analysis utilizing this protocol on a rank-ordering of the response values for each subject is presented in Figure 1. As can be seen from the figure, 70% probability of accurate assessment was achieved for a total of 70% of the responses assessed. It may be interesting to note that the curve also indicates that the sum of the percentages of assessment accuracy and responses assessed remains essentially constant at a value of 140 over the range of 50% to 90% of assessed responses.

2. GROUP A-2

This group consisted of 13 subjects who underwent a test similar to the previous subgroup. On the second repetition each subject was physically attached to a dummy electric shock device and was told that when he lied, a mild electric shock would be administered. It was hoped that this would induce a high stress situation. The equipment did not include any polygraph assessment of stress or deceit.

Assessment of the voice data indicated poor discrimination of the known lies for the group as a whole; hence no detailed statistical analyses were performed. It must be noted that in this particular experiment the time intervals between individual questions and between the questions and their respective elicited responses were of very short duration. As a result numerous subject responses were obscured by the examiner's questions. This rapid-fire questionning was not typical of the majority of examinations processed. In addition, the psychological impact of this very fast questionning is unknown, and hence places real doubt on the validity of the assumption that the psychological forces involved in a typical examination were the same as those involved in this particular experiment.

3. GROUP A-3

Group A-3 consisted of five subjects who underwent interrogation for staged crimes. There was no supportive detailed polygraph evaluation available for these subjects, but subsequent to analysis, information was obtained as to whether each subject was innocent or guilty. There was insufficient information concerning the specifics of the crimes to validate some of the responses, therefore the accuracy of assessment by voice analysis could only be derived for the group as a whole. Using the analysis previously discussed, Figure 2 was developed from the voice response data for this group.

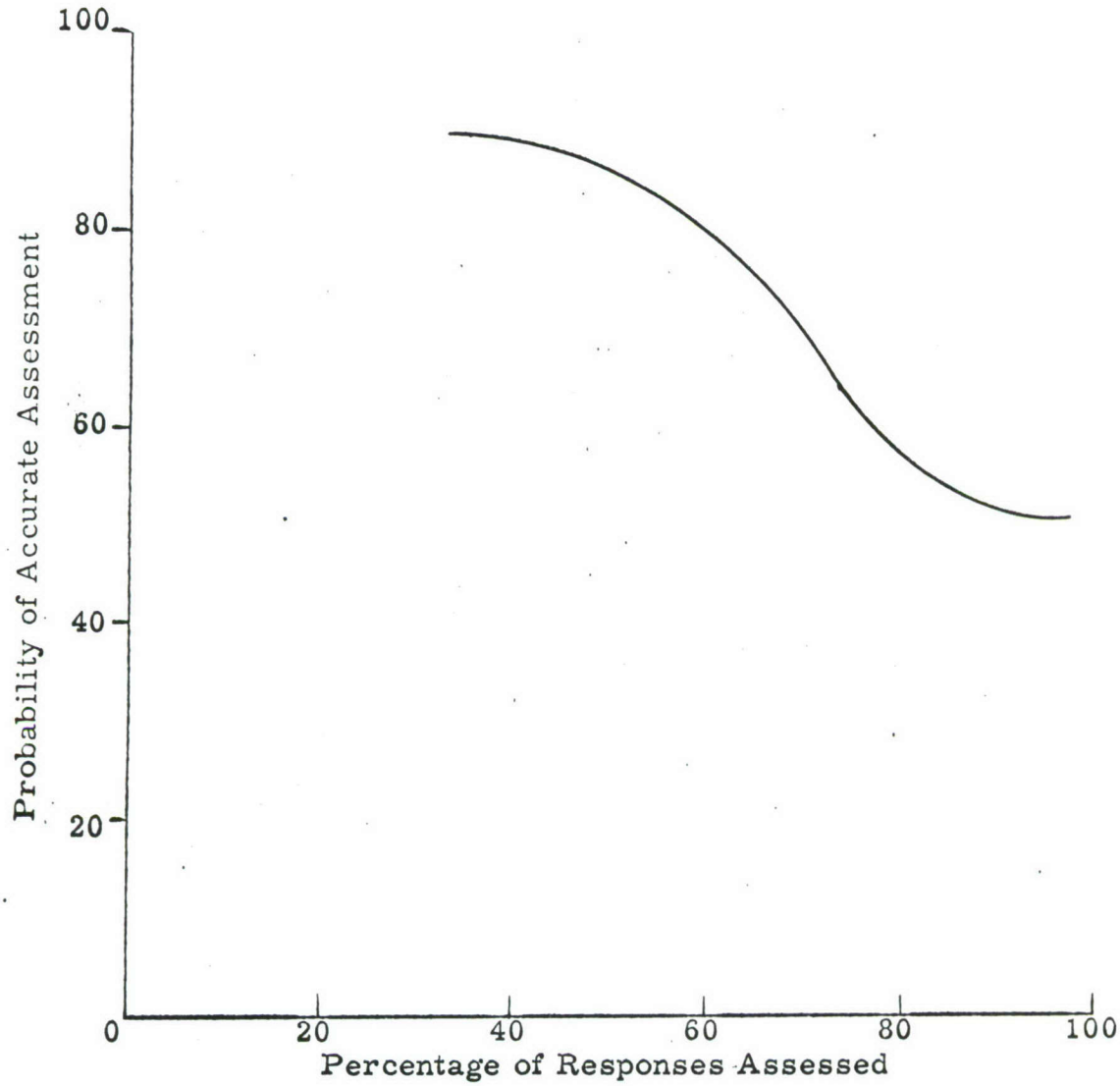


FIGURE I
Accuracy of Assessment versus Percentage of Responses Assessed for Group A-1

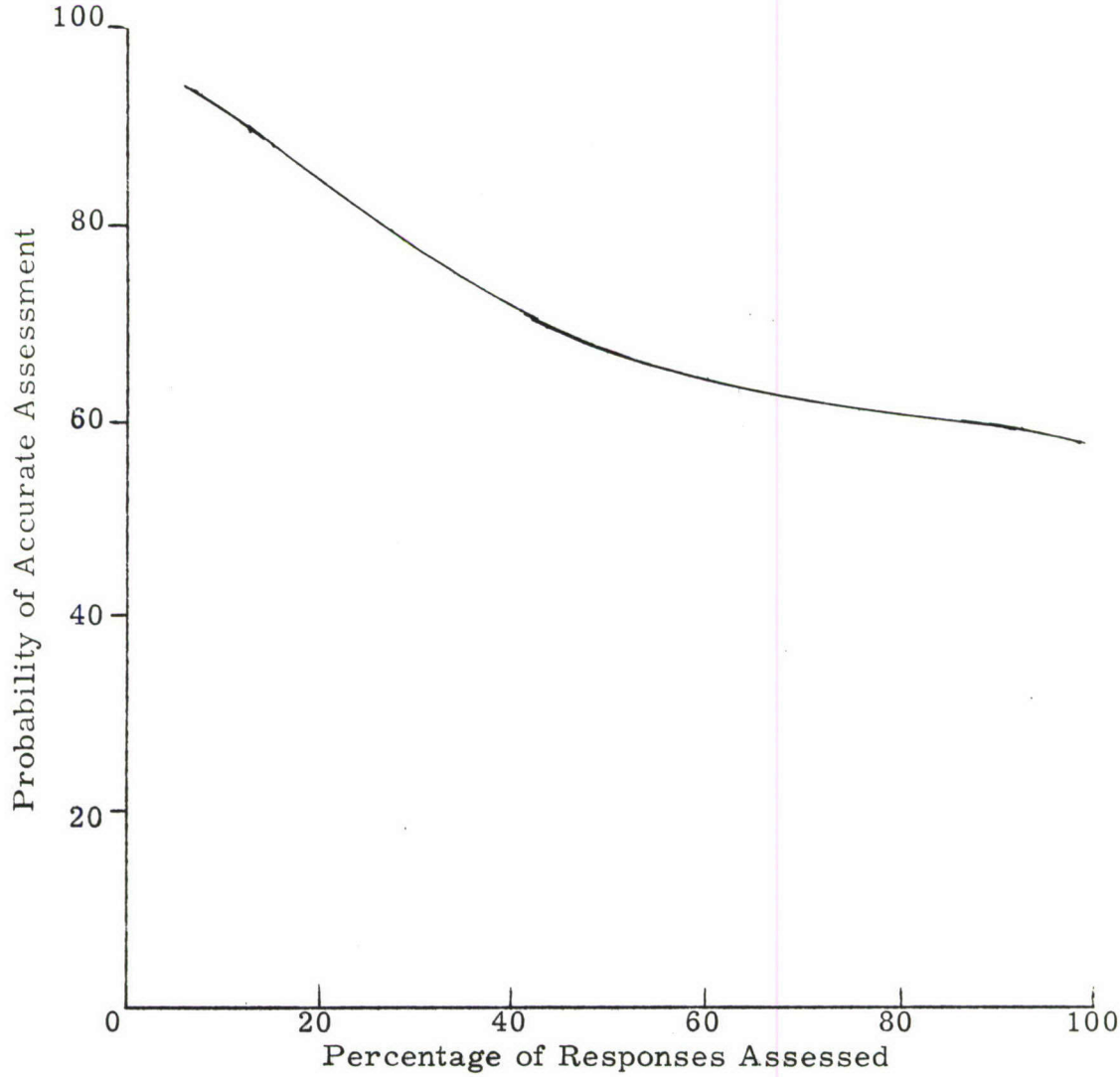


FIGURE 2
Accuracy of Assessment versus Percentage of Responses Assessed for Group A-3

4. GROUP B-1

(1) Discussion of Tests Performed

This group consisted of five subjects examined for actual crimes. These subjects were examined under a fixed protocol in which the same questions were asked in the same sequence in three separate tests (numbered 1, 3, and 4). After the first test, a card stimulation test (numbered 2) was conducted in order to demonstrate to the subject the polygraph's capability to detect deceitful responses. The four tests were performed, with rest periods, within a single session. Three issue questions were asked in Tests 1, 3, and 4 which are intended to relate specifically to the incident under investigation. These were questions numbered 5, 7, and 10 for each subject. Three other stress-inducing questions unrelated to the incident are also asked in Tests 1, 3, and 4. These were questions numbered 4, 6, and 9. An individual is assumed to be deceitful about the issue or incident when a higher value is observed for the ratio measure for the issue questions than for the stress-inducing questions.

The results obtained from the polygraph charts led the evaluators to conclude that Subject W was deceitful, Subjects S and G were not deceitful, Subject B was probably not deceitful while the assessment for Subject A was inconclusive.

The following table summarizes the detailed Polygraph evaluation for two of the subjects in Group B-1.

Subject W (No. 5)

Test Number	Polygraph Evaluation
1	Difficult to evaluate entire group because of equipment problems. However, responses 5 and 7 showed some reaction, with response 7 showing greatest reaction.
2	Responses 4 and 6 showed reaction, with 4 showing greatest reaction with a change occuring in 6.
3	Responses 5 and 7 showed reaction, with 7 showing greatest reaction.
4	Responses 3, 5, 7, and 6 showed stress reactions; their degree of reaction was in the order listed. Responses 3 and 10 also showed some changes.

Subject S (No. 4)

Test Number	Polygraph Evaluation
1	Response 7 showed greatest stress reaction followed by responses 6 and 8.
2	The greatest stress reactions obtained on 2 and 3, with 2 greater.
3	Response 7 showed greatest reaction, closely followed by responses 4 and 5.
4	Same results as Test Number 3.

A detailed polygraph evaluation was unavailable for the remaining three subjects, but the examiner's overall assessments were as follows:

- Subject G (No. 2) - No Deception Indicated (NDI)
- Subject A (No. 3) - Inconclusive
- Subject B (No. 1) - Inconclusive, guilt reactor (probably not deceitful).

The assessment of Inconclusive for Subject B was based principally on the fact that he was determined to be a "guilt reactor". When asked questions relative to a hypothetical crime, this subject reacted in a manner which would normally indicate guilt. Hence, an assessment of DI would be made for this subject, except that his "normal" reaction tends to invalidate such an assessment.

(2) Analysis of Voice Stress Analyzer Results Obtained

It was possible to perform more precise statistical analysis on this group of subjects than for the previous groups since the tests administered provided an unusually consistent and orderly set of data.

A single measure was used to determine whether deception is indicated (i.e., guilt) or not (i.e., innocence) for each of five subjects for which previous assessments had been made. This measure, as discussed previously, is defined as the ratio of the maximum voice signal level at the output of the low bandpass filter (i.e., 100-120 Hz) to that of the high bandpass filter (i.e., 600-800 Hz).

The basic test used for each subject consisted of a series of ten questions, three of which are "crime-specific" or "issue" questions, three are "non-crime" related, but stressful (non-issue), and four are innocuous questions.

The ratio response for the "issue" questions is denoted X_I, while that for the "non-issue" questions is designated X_N. The criteria selected for determining guilt or innocence for each subject is the difference between $\bar{X}_I$ and $\bar{X}_N$ where X_I is the average of the $\bar{X}_I$'s and $\bar{X}_N$ the average value of the X_N's. If X_I is significantly larger than X_N, then deception is indicated, while if X_N is significantly larger than X_I, the reverse is true and no deception is indicated.

The subject was asked the same ten questions in the same order three times with a short break between each of the two repetitions. Only the second and third repetitive "tests" were used in the analysis, giving six X_I and six X_N values for each of the five subjects. The values were first placed in rank order and the X_I's compared against the corresponding X_N's. The average ($\bar{X}_I$, $\bar{X}_N$) and variance ($\sigma^2 X_I$, $\sigma^2 X_N$) were then calculated. The results are shown in Table 2. Note that in three of the five subjects the six values of X_I are either all larger or all smaller than the corresponding X_N values. In one case (Subject No. 4) five of the six X_N's are larger than the corresponding X_I's; while in one subject, all three "usable" X_N's are larger than the corresponding X_I's.

To determine the significance of the difference between the X_I's and X_N's, the following standard statistical test was applied:

(1) Find the difference between the number of plus and minus signs (in each rank-ordered set).

(2) Find the sum of the number of plus and minus signs.

(3) If the difference between the number of plus and minus signs found in (1) is greater than twice the square root of the sum found in (2), then the difference between $\bar{X}_I$ and $\bar{X}_N$ is significant at the one-in-twenty (i.e., 95% confidence) level.

Applying this test to the data presented in Table 2, for the three subjects numbered 1, 2, and 5, we have the following:

(1) Difference d, between number of plus and minus signs = $d = 6 - 0 = 6$.

Ratio Values	Subject Number														
	1			2			3			4			5		
	x_i	x_n	$x_i - x_n$	x_i	x_n	$x_i - x_n$	x_i	x_n	$x_i - x_n$	x_i	x_n	$x_i - x_n$	x_i	x_n	$x_i - x_n$
1	0.95	0.84	+	2.16	4.00	-	0.10	0.10	*	0.80	0.76	+	0.79	0.33	+
2	1.37	0.87	+	2.53	5.70	-	0.13	0.13	*	0.88	0.96	-	1.04	0.52	+
3	1.38	0.90	+	2.84	6.35	-	0.15	0.31	-	0.93	1.10	-	1.78	0.77	+
4	1.92	0.95	+	3.88	12.4	-	0.18	0.35	-	0.95	1.22	-	2.07	1.30	+
5	2.02	1.18	+	9.00	12.9	-	0.25	0.53	-	1.02	2.02	-	2.14	2.00	+
6	2.06	1.58	+	11.0	14.6	-	0.91	0.85	*	1.09	2.56	-	6.17	2.85	+
Σx	9.70	6.32	–	32.2	55.9	–	0.71	1.32	–	5.67	8.62	–	14.0	7.77	–
$\bar{x}$	1.62	1.05	–	5.37	9.31	–	0.18	0.33	–	0.95	1.44	–	2.33	1.29	–
σx	0.45	0.27	–	4.03	4.47	–	.058	.164	–	0.10	0.54	–	1.90	0.97	–

* These values were not used in the statistical analysis (see Discussion).

Table 2

Observed Values for x_i and x_n for 5 Subjects

(2) Sum s, of plus and minus signs: $s = 6 + 0 = 6$;

(3) $2\sqrt{s} = 2\sqrt{6} = 2 \times 2.45 = 4.9$;

(4) $d - 2\sqrt{s} = 6 - 4.9 = 1.1$.

Since $d - 2\sqrt{s}$ is greater than zero, the difference between $\bar{X}_I$ and $\bar{X}_N$ is significant at the 95% confidence level.

For subject number 4, $d - 2\sqrt{s} = 4 - 4.9 = -0.9$, which implies that the difference between X_I and X_N is not significant at the 95% level. However, since we have five minus signs as against only one plus sign, there clearly is a sufficient difference to make us suspect that more information might well have revealed a significant difference.

The total sample size necessary to confirm a significant difference is given by the formula $(2s/d)^2$, or in this case $(2x6/4)^2 = 9$. Hence, three more observations of X_I and X_N are required to establish a significant difference at the 95% confidence level. As it stands, with only six observations, the confidence level is approximately 85%.

For subject number 3, our effective sample size was reduced to only three "useful" observations of X_I and X_N for the following two reasons:

(1) The highest observed value of X_I (i.e., 0.91) is more than three standard deviations above the mean of the sample, and hence statistically does not "belong" to the group.

(2) The lowest observed values for both X_I and X_N (i.e., 0.10 and 0.13) appear to have been below the useful operating range of the equipment, and hence reflect primarily noise as opposed to a true signal output.

As a consequence, using the minimum "total sample size" formula noted above, we need a total of $(2s/d)^2 = (2x3/3)^2 = 4$, or one more observation to establish a significant difference at the 95% confidence level. As it stands however, with only three "useful" observations, the confidence level is about 85%.

The above discussion of the results from the five subjects has dealt only with each subject individually. In order to make an overall evaluation and establish appropriate criteria for determining guilt or

innocence, the magnitude of the ratio function $\bar{X}_I/\bar{X}_N$ or $\bar{X}_N/\bar{X}_I$ was calculated for each subject; the particular ratio chosen in each case being the one that gives a value greater than one. The hypothesis under test is the following: if $\bar{X}_I$ is significantly larger than $\bar{X}_N$, then deception is indicated (guilt), while if X_N is significantly larger than X_I, the reverse is true, indicating innocence.

Table 3 presents the results of these ratios for the five subjects. The results for this group can be summarized as follows:

(1) The ratios $\bar{X}_I/\bar{X}_N$, or $\bar{X}_N/\bar{X}_I$ for the given subjects were all within the narrow range from 1.52 to 1.83. This indicates that from this small sample statistically we can expect, with 95% confidence, that the ratio should fall within the range from 1.4 to 2.0 for a guilt or innocent assessment. It may turn out, however, as a result of further testing, that this range can be extended. However, for the present, if the ratio falls below 1.4 an assessment of inconclusive is indicated.

(2) For subjects numbered 2, 3, 4 $\bar{X}_N$ was greater than $\bar{X}_I$, indicating innocence. This agrees with the previous independent assessment, except for subject #3, the previous assessment was inconclusive.

(3) For subjects numbered 1 and 5, X_I was greater than X_N, indicating guilt. This also agrees with the previous except that for #1 the previous assessment was given as inconclusive by the examiner because the subject appeared to exhibit the characteristics of a "guilt reactor".

5. GROUP B-2

This group consisted of seven individuals suspected of and tested relative to some actual criminal offense. Meetings were held with the polygraph expert who examined all seven subjects and his assessment as to which responses were stressful and indicative of deceit was obtained. The analysis discussed in Section IV was performed to determine the accuracy with which these stressful responses could be discriminated. Figure 3 shows the results of the analysis for this group.

Table 3

Determination of Guilt or Innocence of Five Test Subjects by Means of the Ratios $\bar{X}_I / \bar{X}_N$ or $\bar{X}_N / \bar{X}_I$

Subject No.	$\bar{X}_I$	$\bar{X}_N$	X_I/X_N or $\bar{X}_N/\bar{X}_I$	Assessment (using prototype voice stress analyzer)	Previous Assessment (by military)
1	1.62	1.05	1.54	Guilty	Inconclusive-guilt reaction
2	5.37	9.31	1.73	Innocent	Innocent
3	0.18	0.33	1.83	Innocent	Inconclusive
4	0.95	1.44	1.52	Innocent	Innocent
5	2.33	1.29	1.81	Guilty	Guilty

Average Value	1.69
Standard Deviation	0.15
Range of $\bar{X}_I/\bar{X}_N$ for 95% confidence level	1.4 to 2.0

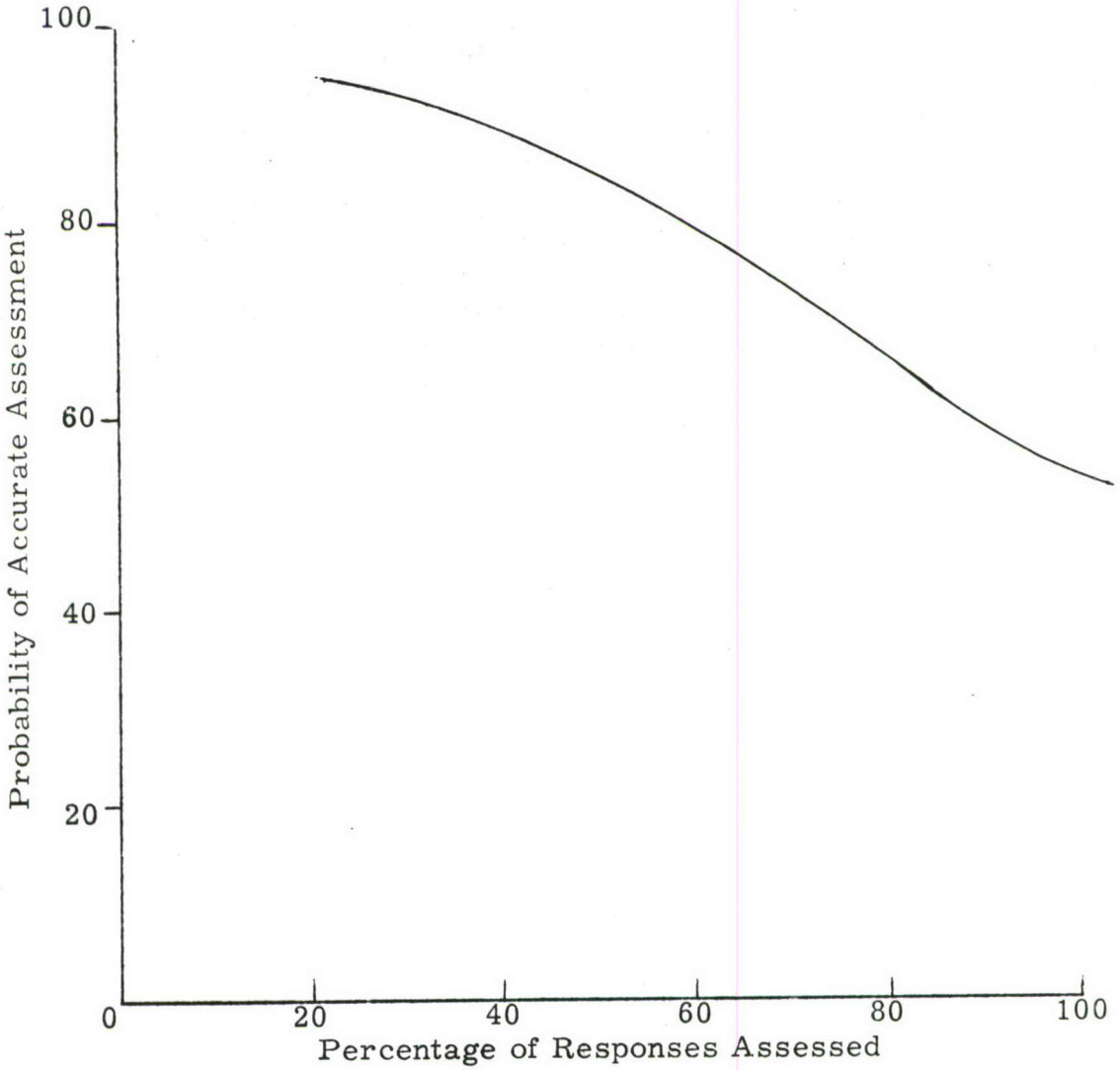


FIGURE 3
Accuracy of Assessment versus Percentage of Responses Assessed for Group B-2

6. GROUP B-3

This group consisted of nine Americans undergoing a laboratory examination for a staged crime. There were three possible roles that each could have played with respect to the crime: innocent, guilty knowledge, or guilty. The roles were selected by dice rolls. The objective of the analysis was to identify the role played by each subject. With an analysis similar to the zone of comparison logic, only three out of the nine were correctly assessed, which is no better than chance. Using the form of analysis developed for the stimulation test produced results which were no better. The only conclusion that can be drawn from the results of the analysis of this group is that the protocol used for this experiment is apparently not amenable to the methodology of voice analysis, as it has thus far developed. It should however be noted that there was considerable distortion and noise present on this group of tapes, which may have had a very significant effect on the voice responses processed.

VI. VOICE STRESS ANALYZER PROTOTYPE

A prototype voice stress analyzer which automatically performs data reduction and records the measure values necessary for assessment of stress is a principal output of this contract. It was initially intended that the breadboard voice stress analyzer previously developed under Contract DAAD05-70-C-0176 would be modified as required to provide the capabilities specified under this contract. However, it was determined that the human engineering requirements called for under the contract were incompatible with the design of this previous voice stress analyzer; hence, a new prototype voice stress analyzer which could satisfactorily meet the contract requirements was designed and built. This new prototype is a moderately complex instrument which performs the basic objectives for which it was developed.

1. PROTOTYPE PERFORMANCE GOALS

The performance goals of the prototype voice analyzer are:

(1) Automatic generation and display of the ratio of the peak values of a voice signal in two different bandpass frequency ranges.

(2) Operation from a tape recorder or microphone input.

(3) Automatic recording of values necessary for decision-making.

(4) Reproducibility of measured outputs to within $\pm$10%.

2. PROTOTYPE PERFORMANCE REQUIREMENTS

The data extraction of the voice analyzer is intended to replicate the technique which had been developed and instrumented in a previously developed engineering model. The data extraction was to have the following signal processing:

(1) The device is to receive a speech signal from either a tape recorder or microphone and divide this signal into two equal value signals.

(2) The device is to filter each of the divided signals. One signal is to be filtered for a bandpass frequency range equivalent to the typical fundamental frequency of human speech, between 80 and 200 Hz. The other of the divided signals is to be filtered for a bandpass frequency range which is experimentally found to be the most appropriate for the voice analysis process.

(3) These bandpass filtered signals are then to have all negative values cut off and be smoothed or partially integrated such that they do not exceed a value of 100 Hz.

(4) The device is then to display the values representative of each and display the ratio values in such a way that a decision can be made as to the stressfulness of the individual response.

3. PROTOTYPE DESIGN

This section contains the schematic drawings (Figures 4, 5, 6 and 7) of the prototype cricuits and equipment and describes the signal flow through these circuits.

The output of a tape recorder is fed into an input connector or, alternatively, a voice signal is received by a microphone and is passed through the microphone input jack into a preamplifier. The voice signal is then taken from the input connector or the microphone preamplifier and passed through a variable amplifier which serves as the primary GAIN control for the unit. The amplified signal is then divided into two equal value signals. One of these signals is filtered for a bandpass frequency range of 80 to 135 hertz. The other is filtered for the bandpass frequency range of 640 to 1080 Hz. These signals are known as the low and high band signals, respectively.

The processing of the two filtered speech signals is exactly the same up to the development of the ratio value, therefore only one of the processing circuits will be discussed. The signal is then passed through a half-wave rectifier and the negative values of the signal are cut off. The signal then enters a capacitor shaping circuit, which is controlled by the LO or HI-SHAPING controls on the panel. Here the signal is partially integrated or smoothed to that extent necessary to dampen background noise accompanying the voice signal. The signal then enters another variable amplifier circuit which is controlled by the LO or HI-GAIN controls on the panel. After gain adjustment, the signal is processed in a "peak detect" circuit. This circuit is used to achieve a varying D.C. level for the voice signal which is easy to read on the meter or chart recorder. The "peak detect" capacitor is normally shorted to ground. This circuit is closed by means of the TALK button or remote talk button. The signal which is permitted to pass through this circuit

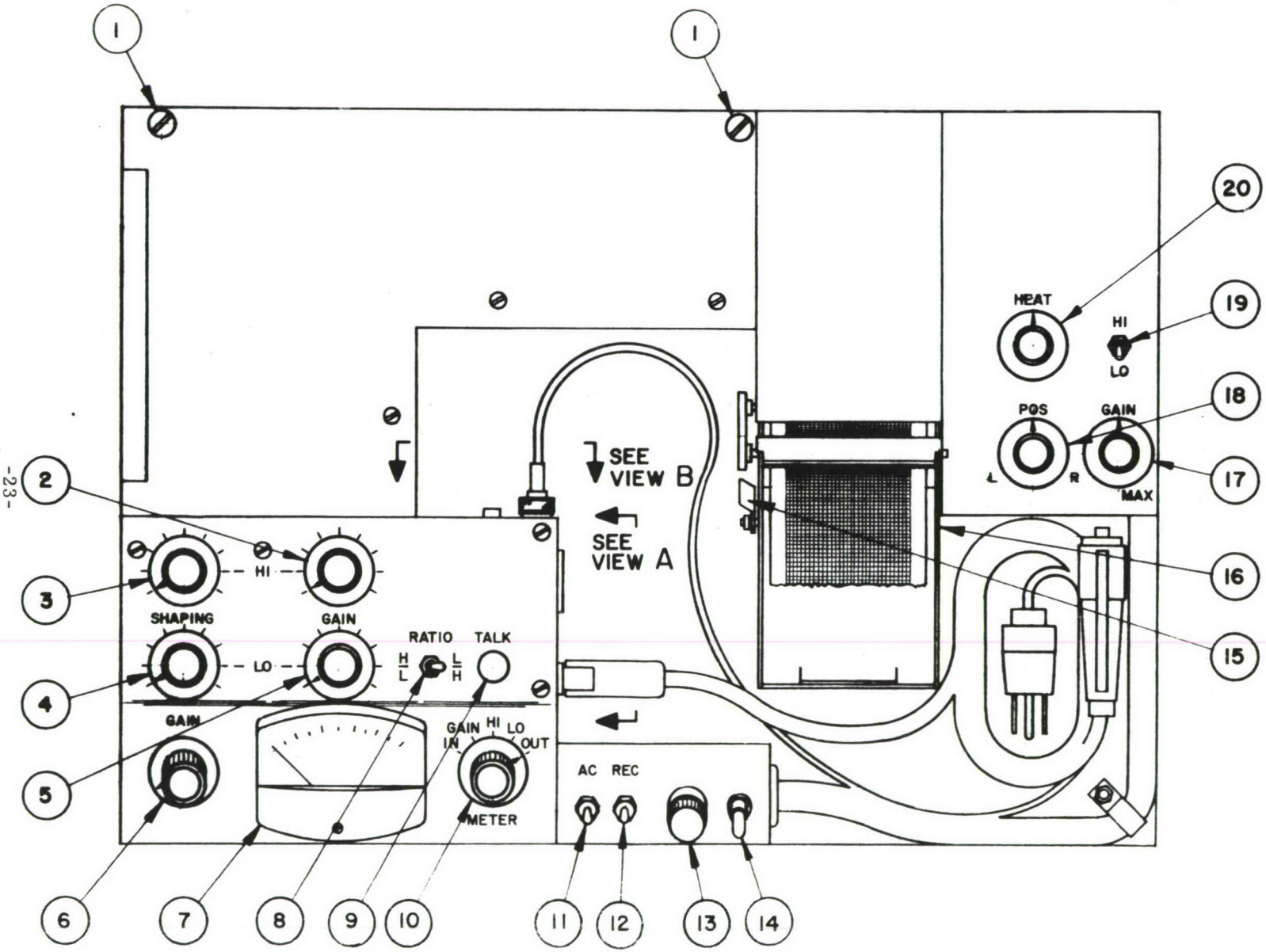
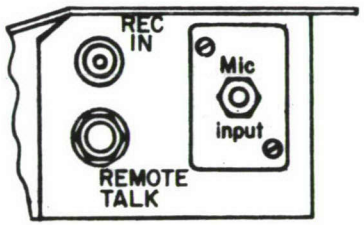


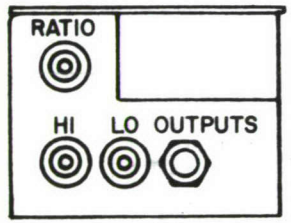
FIGURE 4 - Voice Analyzer Prototype

FIGURE 4 DESCRIPTORS

1. Electronics Access Screws
2. Hi Gain Control
3. Hi Shaping Control
4. Lo Shaping Control
5. Lo Gain Control
6. Primary Gain Control
7. Meter
8. Ratio Select Switch
9. Talk Button
10. Meter Select Switch
11. Master Power Switch
12. Chart Recorder Power Switch
13. Fuse
14. On-Off Lamp
15. Chart Magazine Release Latch
16. Chart Recorder
17. Chart Gain Control
18. Stylus Position Control
19. Paper Drive Speed Control
20. Stylus Heat Control



VIEW A



VIEW B

FIGURE 5
Voice Analyzer Input and Output Connectors

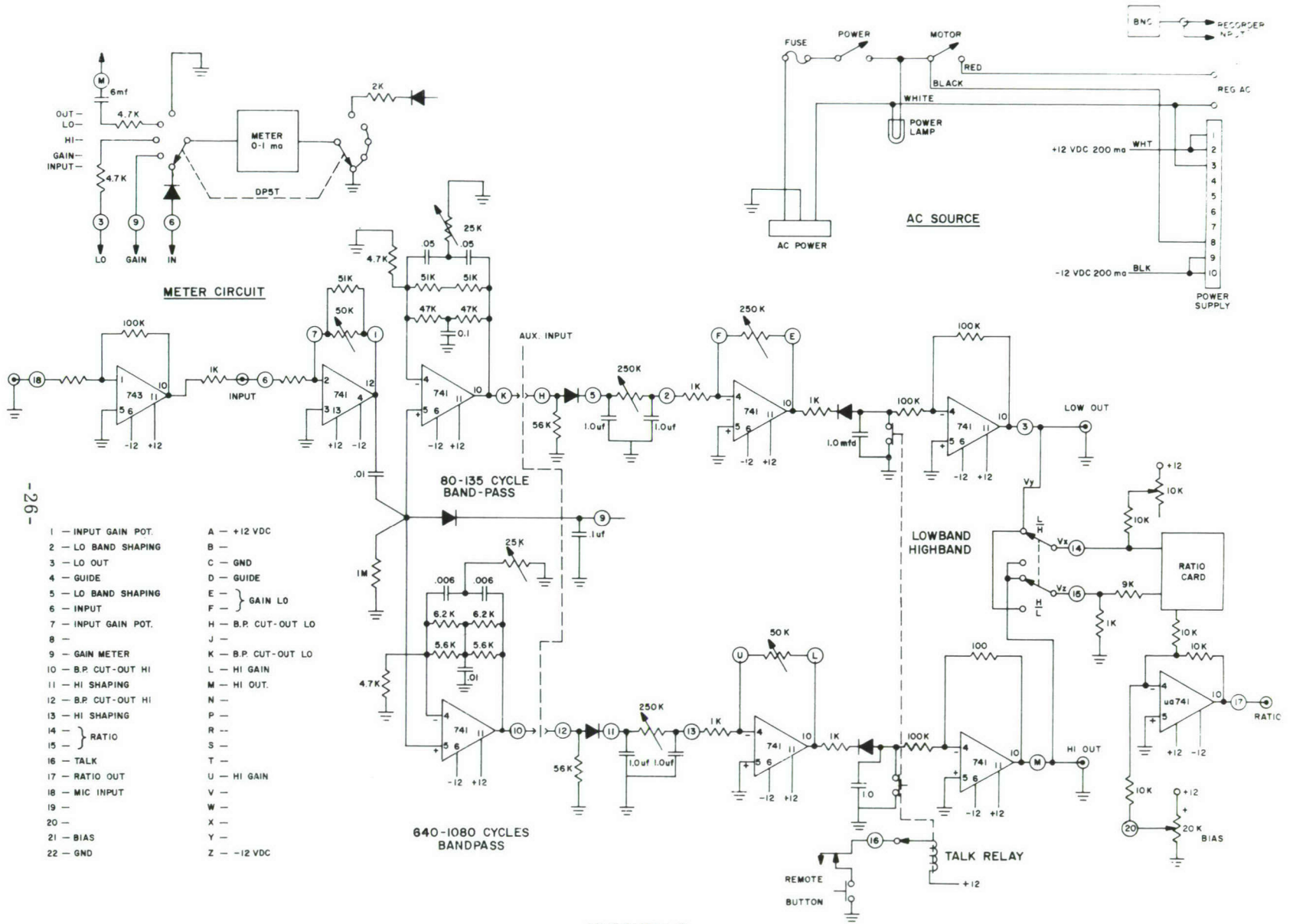


FIGURE 6
Electronic Schematic of Voice Analysis Prototype

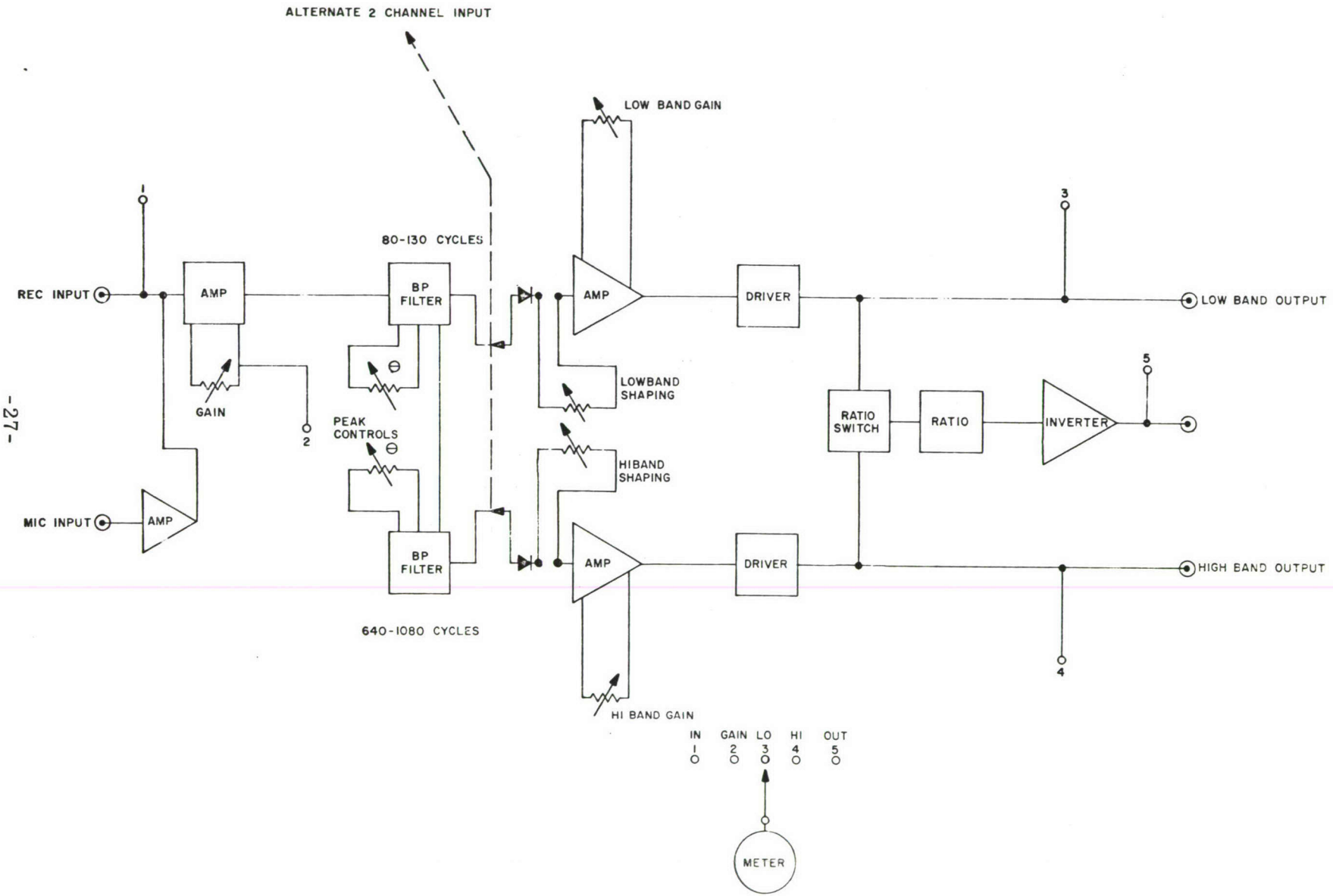


FIGURE 7
Logic Schematic for Voice Analyzer Prototype

goes through a final drive circuit and then put out onto the chart recorder in the low or high modes or is passed into the ratio circuit. The ratio L/H or H/L is selected with the toggle switch controlling the ratio mode. The ratio value is established with the ratio circuit and then put out onto the chart recorder.

The chart recorder is activated by a separate power switch. The chart paper drive is activated by the TALK/remote talk button. There are four controls for the chart recorder: a stylus position control, a stylus HEAT control, a chart drive speed select switch, and a chart GAIN control. These are all connected to the left terminal board of the chart recorder. The GAIN control is used to adjust the signal strength as it is put out on the chart recorder.

The meter which is used to monitor the speech signal is controlled by a five position switch with terminals as indicated in Figures 4 and 7.

4. PROTOTYPE PERFORMANCE RESULTS

The prototype was tested by using the tape recording of Subject G in Group B-1. This tape was run three times. The purpose of the tests was to determine: (1) the reliability of the processing, and (2) the extent to which the prototype ratio outputs corresponded to manual calculations. The prototype was tested for reliability in the low and high band modes, and for computational accuracy in the ratio mode.

The results obtained from these tests showed that the prototype has a reliability on repeated runs of $\pm 10\%$, and provides ratio values which results in the same interpretation as found using manual computation. The test-retest reliability was established by processing the same tape recording through the prototype three times under identical conditions. The output recorded on the chart recorder was manually reduced to measure values and a reliability coefficient computed.

VII. INTERROGATION AND ASSESSMENT PROTOCOL

1. INTERROGATION

The interrogation procedure to be followed in the use of the prototype voice analyzer employs the zone of comparison technique, common to the polygraph. Preceding the actual interrogation, there should be a period of general discussion between the subject and the examiner. This session serves several purposes. First, it is an attempt to get the subject to relax. He is generally in an uncomfortable situation, and therefore likely to be tense. The resulting stress could make the detection of stress as an indication of deceit a more difficult process.

The examiner should attempt to develop a level of rapport with the subject during this session. This will serve as a guide to the examiner in evaluating the stress levels during the interrogation. This session can also be used to determine specific topics which can be used in the questioning. This will be further discussed later, but general areas of the subject's personal life and service career should be discussed. When the examiner feels that the subject is suitably relaxed, he should begin the interrogation.

During the interrogation, the same set of questions should be repeated three times with a short stimulation test immediately following the first set. Within each of the three main sets, there should be three types of questions: (1) innocuous, (2) crime-related, and (3) non-crime-related, but stress-inducing questions. The innocuous questions should be non-stress inducing, hence they should avoid mentioning colors, items of personal attire, matters of a personal nature, or anything which might be related to the crime in question. The best source for the innocuous questions would be the pre-interrogation discussion. For instance, the subject might have stated that he has been in the Service for two years and it appears that he has no negative attitudes concerning this. The question might then be asked, "Have you served in the Army for two years?". Or the subject might have stated that he has a hobby of making cabinetry. He could then be asked, "Do you enjoy making cabinets?". These questions should be used only when the subject has displayed a positive or pleasurable attitude towards the questioned situation.

The crime-related questions should be the simplest to develop. "On March 15, 1972, did you commit the act of_____?", or "Did you steal that money?", et cetera. It is simply a matter of strongly relating the crime and the subject together in one question. The non-crime related, but

stress-inducing questions are a little more difficult, but certain general questions should serve in most cases. They will typically relate to wrongful acts committed during adolescence. "Did you ever steal money from your parents?", "While you were in high school, did you ever cheat on a test?", or possibly, "Have you ever wanted to commit adultery?". There are many such questions which can be developed based on acts that most people have probably committed and yet will still deny. Obviously, these questions should be developed or modified based on information acquired in the discussion session. These questions should not be specifically asked during this discussion as they may lose their stress-inducing value.

Once the questions have been developed, they must be asked in a specific sequence. Four sets of questions are prepared in which the first, third, and fourth set consist of exactly the same questions in the same order. The second set consists of what is termed a stimulation test. The examiner should ask no more than ten to fifteen questions in a set. The sequencing of the questions is most important. The first two or three should be innocuous. These should be followed by alternating crime-related and non-crime-related but stress-inducing questions, with the crime-related question coming first. It is suggested that after the second pair of these, an innocuous question might be asked, followed by a third pair of crime, non-crime question. Each question should be asked in the same order during each repetition.

The stimulation test which follows the first question set should be a standard card stimulation test. The subject should be asked to respond "no" each and every time he is asked if the number he picked was ___. The examiner should ask, "Did you pick number 1?, did you pick number 2?", etc. There should be a pause of approximately three seconds between each response and the next question. During the interrogation the subject must be instructed not to respond immediately after the question is asked. If this is not done, the subject's responses may be obscured by the examiner's question. There must be a pause of at least one second. Caution must also be exercised to avoid extraneous noises which might obscure or distract the subject's responses. Tapping pencils, creaking chairs, and slamming doors are especially interfering. Soundproof rooms are generally unobtainable, but minimal background noise is a must.

2. ASSESSMENT

The assessment consists of the examination of the processed voice signal which has been put out on the chart recorder. After each set of questions has been completed, the examiner should tear off the strip chart and examine the measured peak values of each response. The chart has fifty divisions from side to side with a dark line every five divisions. In the event that the differences in measure value of the responses are not all easily

discernible, the examiner may wish to quickly count up the number of divisions for each response after each test and record the number next to that response peak.

The assessment protocol basically follows what is termed the Zone of Comparison Logic. The first several innocuous questions establish a non-stress pattern, even though the very first response can be interpreted as stressful. The paired crime, non-crime questions should result in a definite pattern of deviation from the previously established non-stress pattern. Empirical evidence has shown that this deviation can occur as either an increase or a decrease in the gross amplitude or height of the responses. Within the deviation from the non-stress pattern the assessment of deceit should be made on the overall deviation of the crime-related responses and the non-crime, stressful responses. This is the zone of comparison. This assessment should be based on the second and third repetitions of the set of questions. It was found that the first repetition was either ambiguous or a reverse of the patterns of deviation of the second and third repetitions. The precise causes are unknown, although it is felt that the stimulation test which follows the first test is a definite factor in bringing out the valid response patterns.

The assessment of the stimulation test is slightly different from the standard peak of tension logic used in polygraph work. The first response acts as a guide in assessing stress. If this response is of a low amplitude relative to the other five, then the lowest or next to lowest (if the first response is the lowest) in amplitude response should be interpreted as the deceitful response. Similarly, if the first response is of a high amplitude relative to the rest, then it is the greatest amplitude which is considered indicative of deceit. If the first response is neither, then the response of greatest deviation is indicative.

The following is a sample set of questions which might be used in an examination:

(1) Is your last name Smith?

(2) Are you stationed at Fort Z?

(3) Do you enjoy classical music?

(4) On April 5, 1972, did you steal three-hundred dollars from the main commissary?

(5) Did you ever steal money from your mother's purse?

(6) Did you steal that three-hundred dollars?

(7) While you were in high school, did you ever cheat on a test?

(8) Do you believe in Santa Claus?

(9) Do you know for sure who stole the money from the commissary?

(10) Have you ever lied to a superior officer?

This section describes the assessment of the prototype analysis of a tape recording made during the interrogation of a subject suspected of theft. Figure 8 is the graphical output of the prototype for this subject's voice responses.

Responses numbered 4, 6, and 9 are responses to non-issue stress-inducing questions. These are X_N's. Responses numbered 5, 7 and 10 are responses to issue questions. These are the X_I's. The sums of the X_I's and X_N's are 69 and 96 respectively. Since the sum of X_I's is less than the sum of X_N's, an assessment of innocent is indicated. To establish a level of confidence for this assessment:

$$X_N/X_I = 96/69 = 1.39$$

which is just outside the experimentally developed lower range of 1.4. Without strong reason to suspect guilt, this assessment of innocent should be accepted.

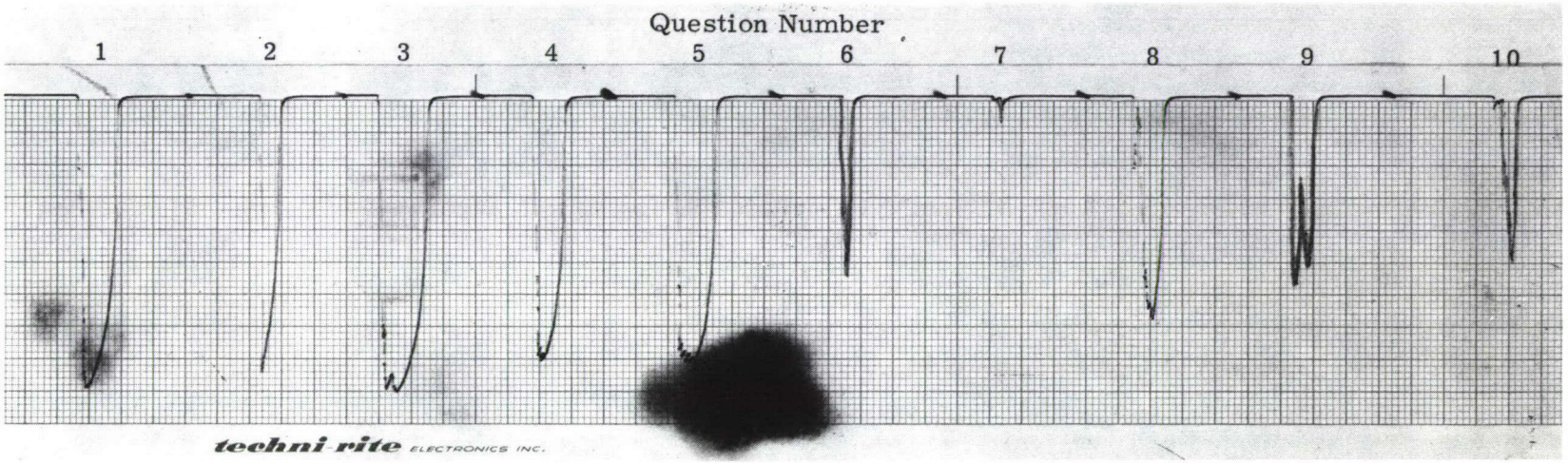


FIGURE 8
Sample Output of Prototype In Ratio Mode

DISTRIBUTION LIST

	Copies
Director Advanced Research Projects Agency Office, Director of Defense Research & Engineering ATTN: Assistant Director, Remote Area Conflict Washington, DC 20310	3
HQDA (DARD-DDS) WASH DC 20310	4
HQDA (DARD-ARZ-C) WASH DC 20310	1
HQDA (DAFD-ZB) WASH DC 20310	1
HQDA (FDCT) WASH DC 20310	2
Commanding General US Army Materiel Command ATTN: Director of Research and Laboratories Washington, DC 20315	1
Commanding General US Army Materiel Command ATTN: Director of Developments Washington, DC 20315	3
Commanding General US Army Materiel Command ATTN: AMCRD-PT Washington, DC 20310	1
US Army Combat Developments Command Liaison Officer Aberdeen Proving Ground, MD 21005	1
Commanding General US Army Combat Developments Command Combat Support Group Fort Belvoir, VA 22060	1
Commanding General US Army Test and Evaluation Command Aberdeen Proving Ground, MD 21005	1

	Copies
Commanding Officer US Army Combat Developments Command Institute of Strategic and Stability Operations Fort Bragg, NC 28307	1
Commanding General US Army John F. Kennedy Center for Special Warfare Fort Bragg, NC 28307	1
Commanding Officer US Army Concept Team in Vietnam APO San Francisco 96384	2
Senior Representative US Army Standardization Group, Australia c/o American Embassy Canberra, A.C.T., Australia	1
Director Air University Library ATTN: AUL3t-64-572 Maxwell Air Force Base, AL 36112	1
Director OSD/ARPA Research & Development Field Unit APO San Francisco 96243	1
Director OSD/ARPA Research & Development Center APO San Francisco 96346	1
Director of Defense Research & Engineering ATTN: Deputy Director for SEA Matters Department of Defense Washington, DC 20310	1
Battelle Memorial Institute Remote Area Conflict Information Center Columbus Laboratories 505 King Avenue Columbus, OH 43201	1
Defense Documentation Center (ASTIA) Cameron Station Alexandria, VA 22314	12

	Copies
Commanding Officer US Army Land Warfare Laboratory ATTN: RDLW-POA Aberdeen Proving Ground, MD 21005	4
Commanding Officer US Army Edgewood Arsenal ATTN: SMUEA-TS-L Edgewood Arsenal, MD 21010	1
USALWL Liaison Officer US Army Concept Team in Vietnam APO San Francisco 96384	1
HQDA (DAMI-ZD) WASH DC 20310	1
US Marine Corps Liaison Officer Aberdeen Proving Ground, MD 21005	1
HQDA (DAMO-PLW) WASH DC 20310	1
HQDA (DAMO-IAM) WASH DC 20310	1

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

DECISION CONTROL, INC. , 4809 Auburn Avenue, Bethesda, Maryland 20014

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

Detection of Emotional Stress by Voice Analysis

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Final Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Fred H. Fuller, Jr.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
July, 1972	33	5

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
DAAD05-71-C-0274	
b. PROJECT NO. LWL C28-71	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Land Warfare Laboratory Aberdeen Proving Grounds Aberdeen, Maryland

13. ABSTRACT

A study was performed by Decision Control Incorporated (DCI) to develop a methodology for determining emotional stress by voice analysis. Tape recordings of 58 subject examinations for actual or staged criminal offenses were electronically processed. The result of this processing was a measure value for the ratio of the amplitudes of the 100-120 Hz to the 600-800 Hz ranges in each subject's spoken response. A methodology for assessing the measure values to determine the relative levels of stress in these responses was developed. A portable prototype voice analyzer was built which electronically processes a voice signal from either a tape recorder or microphone input. An interrogation and assessment protocol similar to the "Zone of Comparison" technique used by many polygraph specialists was developed and used to assess the guilt or innocence of five subjects undergoing "Zone of Comparison" tests. A 95% confidence level was achieved for four of the five subjects.

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Deceit, assessment of by voice analysis Emotional stress, detection of Stress, detection of Voice analysis						